

# **Empfehlungen zur Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten im Kontext erziehungs- und bildungswissenschaftlicher sowie fachdidaktischer Forschung**

## **Gemeinsame Stellungnahme der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE), der Gesellschaft für Empirische Bildungsforschung (GEBF) und der Gesellschaft für Fachdidaktik (GFD) zur Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten in den Erziehungs- und Bildungswissenschaften und Fachdidaktiken**

11. März 2020

### **1 Einleitung**

Die Allianz der Wissenschaftsorganisationen veröffentlichte im Jahr 2010 Grundsätze zum Umgang mit Forschungsdaten<sup>1</sup>. Darin wird die langfristige Archivierung und offene Verfügbarmachung von Forschungsdaten aus öffentlich geförderter Forschung grundsätzlich unterstützt und die Entwicklung von Standards zum Forschungsdatenmanagement empfohlen. Gleichzeitig wird auf die großen Unterschiede zwischen verschiedenen Disziplinen und Datentypen hingewiesen. Auf die Grundsätze der Allianz der Wissenschaftsorganisationen bezugnehmend veröffentlichte die Deutsche Forschungsgemeinschaft (DFG) im Jahr 2015 Leitlinien zum Umgang mit Forschungsdaten<sup>2</sup>, die eine Konkretisierung in Bezug auf von der DFG geförderte Projekte vornehmen. In diesem wird u.a. gefordert, vor Projektbeginn Überlegungen zum Forschungsdatenmanagement anzustellen und im Antrag darzulegen, wie die Forschungsdaten bearbeitet und gesichert und ob und wie sie archiviert werden sollen. Generell sollen, so die DFG, Forschungsdaten langzeitarchiviert und, wenn möglich, zeitnah nach Projektende verfügbar gemacht werden. Die drei Fachgesellschaften unterstützen diese Forderungen grundsätzlich und möchten mit dem vorliegenden Papier einen Beitrag zur Umsetzung dieser Ziele im Feld der erziehungs- und bildungswissenschaftlichen sowie fachdidaktischen Forschung liefern.

Die Fachgesellschaften wurden vonseiten der DFG aufgefordert, Empfehlungen für die Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten ihrer Disziplinen und deren Spezifika zu entwickeln. Daraufhin haben die einzelnen Fachgesellschaften verschiedene Arbeitsschritte unternommen. Die DGfE hatte 2016 eine Arbeitsgruppe etabliert, die sich insbesondere mit den Anforderungen an die Archivierung und Nachnutzung qualitativer Forschungsdaten befasst hat. Die von der Gruppe erarbeitete Stellungnahme zur „Archivierung, Bereitstellung und Nachnutzung qualitativer Forschungsdaten in der Erziehungswissenschaft“<sup>3</sup> wurde 2017 vom DGfE-Vorstand

<sup>1</sup> [https://www.ratswd.de/download/RatSWD\\_WP\\_2010/RatSWD\\_WP\\_156.pdf](https://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_156.pdf); letzter Zugriff: 11.03.2020

<sup>2</sup> [http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien\\_forschungsdaten.pdf](http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf); letzter Zugriff: 11.03.2020

<sup>3</sup> [https://www.dgfe.de/fileadmin/OrdnerRedakteure/Stellungnahmen/2017.09\\_Archivierung\\_qual.\\_Daten.pdf](https://www.dgfe.de/fileadmin/OrdnerRedakteure/Stellungnahmen/2017.09_Archivierung_qual._Daten.pdf) [Zugriff am 02.01.20]. An der Erstellung des Papiers waren beteiligt: Ingrid Gogolin, Helga Kelle, Hans-Christoph Koller, Robert Kreitz, Anja Tervooren und Christine Wiezorek. Der Verbund Forschungsdaten Bildung war durch Doris Bambey und Alexia Meyermann vertreten.

verabschiedet. Die GEBF hat im Jahr 2017 eine Arbeitsgruppe „Forschungsdaten“ zur Erarbeitung eines Empfehlungspapiers eingerichtet. Das von der Arbeitsgruppe erarbeitete Papier „Empfehlungen zum Umgang mit Forschungsdaten“<sup>4</sup> wurde nach Rückmeldungen vom GEBF-Vorstand und weiteren vom Vorstand empfohlenen Expert\*innen von den GEBF-Mitgliedern im Februar 2019 verabschiedet. Die GFD hat das von der GEBF vorgelegte Papier in ihrer 7.Mitgliederversammlung im Februar 2019 mit kleinen Änderungen einstimmig verabschiedet. Um eine gemeinsame Empfehlung der drei Fachgesellschaften zu formulieren, haben sich Vertreter\*innen der drei Fachgesellschaften ausgetauscht und in zwei gemeinsamen Sitzungen die vorliegende Empfehlung konzipiert<sup>5</sup>, die die beiden formulierten Papiere aufgreift und weiterentwickelt. Der konstruktive Dialog der Beteiligten hatte das Ziel, ein Dokument zu verfassen, das nicht nur empfehlenden Charakter für die DFG hat, sondern sich zugleich an die Mitglieder unserer Fachgesellschaften, Begutachtende im Rahmen von Forschungsförderung, weitere Organisationen der Forschungsförderung und an Hochschulen richtet, die sich mit Fragen der *Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten* im Kontext von Forschung, Wissenschaft und Lehre in erziehungs- und bildungswissenschaftlichen sowie fachdidaktischen Kontexten aus unterschiedlichen Perspektiven auseinandersetzen. Die Beschäftigung mit dem Themenfeld beginnt i.d.R. mit der Antragstellung für Forschungsprojekte, die entsprechende Angaben einfordern.

Alle drei formulierten Zielsetzungen – die *Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten* – sind auch im Rahmen von *Open Science relevant*, also einer offenen und öffentlichen Wissenschaft, die danach strebt, Forschung der Fachöffentlichkeit und allgemeinen Öffentlichkeit zugänglich und nachvollziehbar zu machen. Die prinzipielle Möglichkeit der Re-Analyse von Daten oder Replikation von Studien und eine Transparenz des Auswertungsprozesses stellen einen zentralen Qualitätsaspekt empirischer Forschung dar. Diese Prämissen einer offenen Wissenschaft unterstützen die drei Fachgesellschaften ausdrücklich.

Wenngleich die Forschungsfelder der drei Fachgesellschaften DGfE, GEBF und GFD nicht identisch sind, so ist den empirisch arbeitenden Teilbereichen der Fachgesellschaften gemeinsam, dass sie wesentlich durch die Erhebung von Daten konturiert sind, die sich auf Bildung, Erziehung, Lernen, Lehren und Entwicklung von bzw. durch Menschen auszeichnen. Dieser Forschungsgegenstand führt zu diversen Spezifika in der Datenerhebung.

Bildungs-, Erziehungs-, Sozialisations-, Lern-, Lehr- und Entwicklungsprozessen stellen (a) *vulnerable Prozesse* dar, an deren Untersuchung sich besondere Herausforderungen stellen. Unabhängig von den Prozessen stellen (b) Minderjährige, also Kinder und Jugendliche, eine besonders schützenswerte *Gruppe* dar, die zugleich nicht allein entscheiden kann und darf, ob und wie sie sich an Forschungsprozessen beteiligen möchte oder nicht. Zudem erfolgt die Datenerhebung (c) vielfach im *Kontext staatlich organisierter pädagogischer Einrichtungen*, wie z.B. Schule oder sozialpädagogischen Angeboten, in denen sich die Akteur\*innen in sozialen Rollen begegnen, die mit unterschiedlichen Erwartungen an das soziale Miteinander einhergehen und die letztlich hierarchisch organisiert sind. Die drei genannten Aspekte stellen Spezifika erziehungs- und bildungswissenschaftlicher sowie fachdidaktischer Forschung dar, denen gemeinsam ist, dass sie eine besondere Reflexion der datenschutzrechtlichen Bestimmungen als auch der forschungsethischen Implikationen erfordern, die im Rahmen des Dokuments aufgegriffen und diskutiert werden sollen.

---

<sup>4</sup> AG Forschungsdaten im Auftrag des GEBF Vorstands (2019). Empfehlungen zum Umgang mit Forschungsdaten. An der Erstellung des Papiers waren beteiligt: Malte Jansen (Redaktion), Cordula Artelt, Marita Jacob, Jan Marcus, Benjamin Nagengast (Vertretung: Norman Rose), Gabriel Nagy, Beatrice Rammstedt, Petra Stanat, Maik Walpuski als GEBF-Mitglieder sowie Reiner Mauer, Alexia Meyermann (Vertretung: Doris Bambej) als Beratende Mitglieder. Eine ergänzende Kommentierung erfolgte durch Oliver Dickhäuser, Tina Hascher und Susanne Prediger.

<sup>5</sup> Die Fachgesellschaften wurden vertreten durch: Tanja Sturm, Christine Wiezorek und Ingrid Miethe für die DGfE, Malte Jansen, Johannes Hartig und Mareike Kunter für die GEBF und Michael Hemmer und Friederike Korneck für die GFD.

Eine weitere Besonderheit stellt die Vielfalt methodologisch-methodischer Zugänge dar, die in erziehungs- und bildungswissenschaftlicher sowie fachdidaktischer Forschung Anwendung finden. Entsprechend entstehen unterschiedliche Formen von Daten, die ihrerseits je spezifische Formen des Datenschutzes erfordern, mit spezifischen forschungsethischen Fragen verbunden sind, unterschiedliches Nachnutzungspotenzial für sekundäranalytische Forschung aufweisen und mit je unterschiedlichem Arbeitsaufwand, der für die Bereitstellung von Daten erforderlich ist.

Eine weitere Besonderheit der erziehungs- und bildungswissenschaftlichen sowie fachdidaktischen Forschung ist es, dass die Datenerhebung für die Institution und die Teilnehmenden mit teilweise erheblichen zeitlichen Aufwand verbunden ist. Der Feldzugang selbst stellt somit eine besondere Herausforderung unserer Forschung dar. Entsprechend wichtig ist ein verantwortungsvoller Umgang mit den unter diesen Bedingungen erhobenen Daten. Dies gebietet auch eine effiziente Nutzung von für die Erhebung eingesetzten Mitteln. Die Bereitstellung von Forschungsdaten für Sekundäranalysen kann dazu beitragen, dass diese effiziente Nutzung durch die umfassende Auswertung vorhandener Datensätze gewährleistet wird und Teilnehmende nicht durch zusätzliche Erhebungen belastet werden, wenn Fragestellungen auch sekundäranalytisch zu beantworten wären.

Daher sprechen wir uns grundsätzlich dafür aus, dass Forschende prüfen, unter welchen Bedingungen die im Rahmen von Forschungsprojekten zu erhebenden Daten bereitgestellt und nachgenutzt werden können. Die Archivierung, Bereitstellung und Nachnutzung von Daten werfen jedoch auch Fragen nach *Datenschutz*, *Forschungsethik*, der *Autorenschaft* sowie dem *Aufwand* für die Datenaufbereitung auf, aber auch des *Nachnutzungspotenzials der Daten für die wissenschaftliche Community*, die für jedes Projekt im Einzelnen zu überprüfen und zu formulieren sind. Die Bereiche sind miteinander verbunden und auch in ihrer Bedeutung für das primäre Forschungsprojekt und seine Durchführung zu prüfen. Hierzu werden im Folgenden Empfehlungen gegeben, die eine Bewertung des adäquaten Umgangs mit den Forschungsdaten erleichtern sollen.

In Anbetracht der Herausforderungen bei der Datenerhebung wird auch deutlich, dass die an Datenerhebungen beteiligten Primärforschenden ein berechtigtes Interesse an der Honorierung ihrer Erhebungsaktivitäten haben. Ziel der Empfehlungen ist daher immer auch, die Interessen von datenerhebenden Primärforscherinnen und Primärforschern auf der einen Seite und datennutzenden Sekundärforscherinnen und Sekundärforschern auf der anderen Seite gleichermaßen zu berücksichtigen, und eine hohe Wertschätzung der mit Primärerhebungen verbundenen Arbeit zum Ausdruck zu bringen.

## 2 Terminologie

### 2.1 Archivierung, Bereitstellung und Nachnutzung

Unter *Archivierung* wird die Sicherung von Forschungsdaten, meist in digitalisierter Form, verstanden. Den Leitlinien zur Sicherung guter wissenschaftlicher Praxis der DFG folgend sollen Forschungsdaten für mindestens 10 Jahre gesichert werden<sup>6</sup>. Dabei kann die Archivierung in der eigenen Institution oder in einem hierfür geeigneten Repositorium, über die sie recherchiert werden können, erfolgen. *Bereitstellung* bedeutet die Aufbereitung und Dokumentation von Forschungsdaten zum Zweck ihrer Zugänglichkeit für Sekundärforschende. Der Begriff der *Nachnutzung* umfasst Formen der Reanalyse der Forschungsdaten zur Replikation von Forschungsergebnissen oder zur Untersuchung neuer Fragestellungen durch die Sekundäranalyse gespeicherter Forschungsdaten.

---

<sup>6</sup> [https://www.dfg.de/download/pdf/foerderung/rechtliche\\_rahmenbedingungen/gute\\_wissenschaftliche\\_praxis/kodex\\_gwp.pdf](https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf); letzter Zugriff: 10.02.2020

Mit der Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten werden Zielsetzungen verbunden, die teilweise unabhängig voneinander sind. So sollen zum einen Forschungsdaten und die für das Zustandekommen von Forschungsergebnissen relevanten Informationen im Sinne guter wissenschaftlicher Praxis dokumentiert und archiviert werden. Zum anderen soll die sorgfältige Dokumentation und Archivierung von Forschungsdaten dazu dienen, im Rahmen der eigenen Forschungsarbeit weitere Auswertungen (zu einem späteren Zeitpunkt) oder Replikationen von Ergebnissen, auch unabhängig von einer Nachnutzung durch Sekundärforschende, zu ermöglichen. Schließlich kann die Archivierung von Forschungsdaten auch die Möglichkeit einer Nachnutzung durch Sekundärforschende beinhalten.

Die Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten ist aufgrund der Diversität (meta-) theoretischer und methodologischer Zugänge und damit der Gegenstände, Fragestellungen, Erhebungsmethoden und Auswertungs- und Interpretationsverfahren mit sehr unterschiedlichen Herausforderungen *des Datenschutzes, des Urheberrechts, der Forschungsmethodik, Forschungsethik und nicht zuletzt der Forschungsökonomie* konfrontiert. Die Komplexität der damit verbundenen Probleme und Fragen verlangt einen hohen Grad an Sensibilität und Expertise seitens der Primärforschenden, welche die Daten erheben, seitens der Institutionen, die die Daten archivieren und bereitstellen, seitens der Sekundärforschenden, welche die Daten nachnutzen, und nicht zuletzt seitens der Forschungspolitik sowie der Organisationen der Forschungsförderung, die die Forschungsvorhaben (finanziell) unterstützen bzw. ermöglichen. Dennoch sollten Forschungsdaten öffentlich geförderter Projekte für Sekundäranalysen bereitgestellt werden. Die Zugänglichkeit von Daten zur Nachnutzung kann dabei unterschiedlich stark reguliert werden. (siehe Kasten 3 in Abschnitt 4.1).

## 2.2 Arten von Daten

Die Allianz der deutschen Wissenschaftsorganisationen definiert Forschungsdaten wie folgt<sup>7</sup>:

*„Forschungsdaten sind Daten, die im Zuge wissenschaftlicher Vorhaben z. B. durch Digitalisierung, Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen“.*

Zumeist werden Forschungsdaten in quantitative und qualitative Daten unterschieden. Weitere Klassifizierungen beziehen sich auf das Datenformat oder ‚Trägermedium‘ (z.B. numerische, Text-, Bild- und Videodaten.) oder unterscheiden personenbezogene von organisations- oder institutionenbezogenen Daten. In Bezug auf Fragen der Archivierung von Forschungsdaten erscheint es sinnvoll, zunächst nach drei Kategorien von Daten zu differenzieren, die sich auf den Forschungsprozess selbst und die Frage der Aufbereitung von Daten für die Auswertung (z.B. Anonymisierung und Zugänglichmachung) beziehen: (a) unbearbeitete Rohdaten, (b) für Forschung (und Archivierung) aufbereitete Daten sowie (c) Kontextdaten bzw. –informationen.

---

<sup>7</sup> <http://www.allianzinitiative.de/de/archiv/forschungsdaten>; letzter Zugriff: 11.03.2020

#### a) Unbearbeitete Rohdaten

Dies sind unmittelbar aus der Erhebung angefallene Originaldaten, die noch in keinerlei Weise bereinigt, ausgewertet oder anderweitig nachbearbeitet wurden. Diese Daten lassen sich nach Medienformat und Datenquellen unterscheiden, z. B.:

- Audio- oder Videoaufnahmen von Interviews, Gruppendiskussionen, Interaktionen, pädagogischen Situationen im Unterricht etc.
- Artefakte und Dokumente: Briefe, Zeichnungen, Akten, Konzepte, Zeugnisse, Fotos, Schriftstücke etc.
- Antwortdaten aus Test- oder Fragebogenverfahren (auf Papier oder computerbasiert)
- Messdaten aus physiologischen Verfahren (z. B. Herzrate, EEG) oder Eye-Tracking-Untersuchungen
- Log-Daten von Mensch-Computer-Interaktionen (z. B. aus der Bearbeitung computerbasierter Testaufgaben oder dem Verhalten in computerbasierten Lernumgebungen)
- Beobachtungsprotokolle und Feldnotizen

#### b) Aufbereitete Daten

In der Regel werden Daten archiviert, die bereits für eine Auswertung bearbeitet wurden. Relevant in Bezug auf die Archivierung ist die ‚Struktur‘ des aufbereiteten Datenmediums. Interviews können in Form codierter Transkripte aufbereitet werden, Antwortdaten aus Tests und Fragebögen können als *pseudonymisierte oder anonymisierte Datensätze* (Definition siehe II.3) in unterschiedlichen Bearbeitungsstadien (z.B. Bewertung von Antworten, Imputation fehlender Werte) gespeichert und Dokumente als Transkripte, Zeichnungen, (teilweise geschwärzte) Kopien oder Bilder aufbereitet werden.

Dass die Grenze zwischen Rohdaten und aufbereiteten Daten mitunter fließend ist bzw. sich nicht alle (personenbezogenen) Informationen von Rohdaten ohne Verlust für die Auswertung pseudonymisieren lassen, wird anhand bildförmiger Daten wie Video- und Fotografien und anhand von Beobachtungsprotokollen deutlich.

#### c) Kontextdaten/Kontextinformationen

Eine Datenerhebung findet immer in konkreten (institutionellen, organisatorischen oder lebensweltlichen) Kontexten statt und ist durch diese situiert. Als Kontextdaten wird insofern die Gruppe von Daten und Informationen bezeichnet, die Informationen über den Erhebungskontext der ‚Objektdaten‘ enthalten. Kontextdaten sind personen-, orts- und situationsbezogene Angaben, die datenschutzrechtlich besonders geschützt sind, wie z.B. Anonymisierungsregeln und -listen oder Kodierleitfäden. Für viele Fragestellungen können die ‚eigentlichen‘ Daten (als Rohdaten oder in aufbereiteter Form) erst durch diese Informationen sinnvoll ausgewertet werden. Beispiele für Kontextdaten sind allgemeinere Ortsangaben wie Bundesland oder Stadt, spezifischere Orts- oder Personenangaben wie Schulnamen, die Funktionen von Personen in Institutionen oder Informationen zur Erhebungssituation wie Ort und Zeit eines Interviews, aber auch Experten\*innenwissen, das im Rahmen der Forschung eher informell erworben wird sowie ‚Insiderwissen‘, das durch Erfahrungen entsteht, die, bspw. in ethnographischen Forschungskontexten, die/der Primärforschende als Person erworben hat.

Als eine weitere Kategorie von Dokumenten im Kontext der Datenarchivierung kommt den verwendeten Erhebungsinstrumenten eine zentrale Bedeutung zu. Spezifische theoretische Konstrukte oder Modellierungen können z. B. nur mit verschiedenen Test- oder Fragebogenverfahren

operationalisiert werden, während in anderen Interview- oder Gesprächsformaten Fragen offen und ad hoc formuliert werden. Von daher ist eine umfassende Dokumentation der Erhebungsinstrumente für eine sinnvolle Auswertung oft unerlässlich. Wurden eigens entwickelte oder angepasste Instrumente zur Datenerhebung verwendet, sollten diese – als Teil der Kontextdaten/Kontextinformationen - zugänglich gemacht werden. Wurden bereits publizierte Instrumente eingesetzt, liefert eine Zitation derselben möglichen Nachnutzenden die notwendigen Informationen. Insofern sind bei den Erhebungsinstrumenten noch keine Datenschutzfragen, jedoch Fragen des Urheberrechts zu beachten.

### **2.3 Datenschutzrechtliche Konzepte nach § 4 DSGVO und Erwägungsgrund 26 (Stand Januar 2020)**

*Personenbezogene Daten:* Alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden „betroffene Person“) beziehen, sind personenbezogene Daten. Als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen identifiziert werden kann, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind. Für die Speicherung, Archivierung und Weitergabe personenbezogener Daten ist das informierte Einverständnis der Personen erforderlich.

*Verarbeitung:* Als Verarbeitung personenbezogener Daten wird jeder mit oder ohne Hilfe automatisierter Verfahren ausgeführte Vorgang oder jede Vorgangsreihe der Sammlung personenbezogener Daten angesehen. Dazu gehören Erheben, Erfassen, Organisieren, Ordnen, Speicherung, Anpassung oder Veränderung, Auslesen, Abfragen, Verwendung, Offenlegung durch Übermittlung, Verbreitung oder eine andere Form der Bereitstellung, den Abgleich oder die Verknüpfung, die Einschränkung, das Löschen oder die Vernichtung von Daten.

*Pseudonymisierung:* Die Verarbeitung personenbezogener Daten in einer Weise, dass diese ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, wird als Pseudonymisierung bezeichnet. Die zusätzlichen Informationen – Kontextdaten- und Informationen – müssen gesondert aufbewahrt werden, was eigene technische und organisatorische Maßnahmen erfordert. Diese müssen gewährleisten, dass die Informationen nicht einer natürlichen Person zugewiesen werden können.

*Anonymisierung von Daten:* Forschungsdaten ohne Personenbezug. Unter einer Anonymisierung versteht man eine Veränderung der Daten, die dazu führt, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr (sog. absolute Anonymisierung) oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft (sog. faktische Anonymisierung) einer bestimmten oder bestimmaren natürlichen Person zugeordnet werden können. Diese anonymisierten Daten fallen *nicht* unter den Anwendungsbereich des Datenschutzrechts und können somit grundsätzlich unter Berücksichtigung sonstiger eventuell zu beachtender Rechte (z.B. Urheberrechte) frei verarbeitet werden. Datenschutzrechtliche Restriktionen lassen sich somit durch einen sparsamen Gebrauch von personenbezogenen Daten sowie durch eine Anonymisierung der für den Forschungszweck notwendigerweise zu erhebenden personenbezogenen Daten vermeiden.

Forschungsdaten können auf sehr unterschiedliche Weise in ethischer und datenschutzrechtlicher Hinsicht sensibel sein. Während in einigen Fällen mit vergleichsweise geringem Aufwand eine Deanononymisierung der Forschungsdaten praktisch ausgeschlossen werden kann, ist es in anderen Fällen aufgrund der Charakteristik der untersuchten Population und/oder der Art des Umfangs der

erhobenen Informationen über einzelne Personen nicht möglich, die Daten zu anonymisieren, ohne dass sie ihren Informationsgehalt verlieren.

### **3 Praktische Empfehlungen zum Umgang mit Forschungsdaten**

Im Folgenden werden Empfehlungen zum Umgang mit Forschungsdaten formuliert, die alle Phasen der Projektplanung und -durchführung betreffen. Diese Hinweise sollen Forschenden die Planung und Umsetzung ihrer Forschungsprojekte erleichtern sowie Gutachtenden helfen, die Qualität von Forschungsanträgen auch im Hinblick auf die Überlegungen zur Archivierung, Bereitstellung und Nachnutzung von Daten zu bewerten.

#### **3.1 Vor Projektbeginn/im Rahmen der Beantragung von Forschungsprojekten**

Im Rahmen der Planung und/oder Beantragung von Forschungsprojekten, in deren Rahmen neue Daten erhoben werden, sollte ein Forschungsdatenmanagementplan erstellt werden, der Auskunft darüber gibt, welche Rohdaten erhoben werden, wie diese aufbereitet und gespeichert werden, wie sie nach Projektende archiviert werden sollen, welche Dokumentationsmaterialien erstellt werden sollen sowie, ob und in welcher Form sie zur Sekundärnutzung verfügbar gemacht werden (können) (siehe Kasten 1). Das gilt auch für die Kontextdaten und Kontextinformationen, die häufig zur Kontextualisierung bzw. zum Verständnis der aufbereiteten Daten benötigt werden. Der Datenmanagementplan kann „auch begründet vorsehen, dass erst im Projektverlauf und abhängig von den Erfordernissen des Forschungsprozesses konkret über Details des Datenzugangs entschieden wird.“<sup>8</sup> <sup>9</sup> Soweit dies inhaltlich vor Abschluss des Projekts möglich ist, sollten potenzielle Nachnutzungsmöglichkeiten benannt und das Nachnutzungspotenzial der generierten Forschungsdaten abgeschätzt werden. Verknüpft damit sollte im Datenmanagementplan auch dargestellt werden, welche Ressourcen für das Forschungsdatenmanagement während des Projekts eingeplant sind. Im Datenmanagementplan ist aufzuführen, ob und in welcher Form Daten mit direktem Personenbezug erhoben und wie diese gespeichert werden. Dazu empfiehlt sich die Rücksprache mit der bzw. dem Datenschutzbeauftragten und ggf. der Ethikkommission der eigenen Organisation bzw. Fachgesellschaft. Im Fall der Bereitstellung der Daten für Sekundärnutzung ist im Datenmanagementplan darzulegen, wie bzw. wo die Daten archiviert werden. Grundsätzlich gelten die Regeln guter wissenschaftlicher Praxis und sind die jeweiligen datenschutzrechtlichen Regelungen zu berücksichtigen. Die Aufbereitung der Daten zur Archivierung bzw. Bereitstellung soll als eigenständiger Teil der Projektplanung verstanden und mit entsprechenden Ressourcen berücksichtigt werden.

---

<sup>8</sup> RatSWD (2015): Stellungnahme des RatSWD zur Archivierung und Sekundärnutzung von Daten der qualitativen Sozialforschung, S. 7. Online verfügbar unter: [https://www.ratswd.de/dl/RatSWD\\_Stellungnahme\\_QualiDaten.pdf](https://www.ratswd.de/dl/RatSWD_Stellungnahme_QualiDaten.pdf); letzter Zugriff: 26.11.2019

<sup>9</sup> Der RatSWD (2015) argumentiert hier, dass „auf diese Weise die Frage, ob aus forschungspraktischen, datenschutzrechtlichen oder ethischen Gründen Sekundäranalysen nicht möglich oder nicht ratsam sind, transparent innerhalb der geeigneten Scientific Community geklärt“ (ebd.) werden kann.

### **Kasten 1: Checkliste für einen Datenmanagementplan bei der Antragstellung für Forschungsprojekte**

Aussagen zum Umgang mit den zu gewinnenden Forschungsdaten werden von vielen Forschungsförderinstitutionen (z. B. DFG, BMBF) bereits in der Antragstellung erwünscht. Ein solcher Datenmanagementplan sollte folgende Punkte enthalten<sup>10</sup>:

- Datenarten und Datennutzung: Welche Arten von Daten erheben Sie und wie werden sie genutzt?
- Dokumentation: Wie kann Ihre Forschung für andere nachvollziehbar gemacht werden?
- Rechtliche und ethische Aspekte: Datenschutz, Urheberrecht, ethische Aspekte
- Datenspeicherung und –sicherung: Wie werden Daten gespeichert, wie ist der Zugang geregelt?
- Archivierung und Nachnutzung der Forschungsdaten:
  - Aussagen zum Nachnutzungspotenzial (hoch, gering etc.), Gründe für diese Einschätzung
  - Aussagen zu Aspekten, die einer Bereitstellung für die Nachnutzung entgegenstehen
  - Aussagen zur geplanten Bereitstellung für Nachnutzung
- Erforderliche Ressourcen

### **3.2 Vor den Erhebungen: Einverständniserklärungen**

Um Daten erheben zu können, bedarf es zunächst des Einverständnisses der Beforschten bzw. ihrer Erziehungsberechtigten und/oder ihrer rechtlichen Betreuung. Ist die Forschung im Kontext gesellschaftlicher Organisationen und Einrichtungen angesiedelt, z.B. Schule, Kindertageseinrichtung, bedarf es zuvor außerdem einer Genehmigung des Vorhabens durch die zuständigen Träger und/oder Behörden. Erfahrungen der letzten Jahre zeigen, dass nicht selten die Weitergabe der Daten an Dritte von dieser Seite ausgeschlossen wird und Vorgaben gemacht werden, die besagen, dass die Daten nur für das Projekt selbst zu nutzen sind. Entsprechend ist eine mögliche Archivierungsabsicht im Rahmen des Genehmigungsverfahrens explizit zu thematisieren und, soweit möglich, dafür zu plädieren, dass von behördlicher Seite keine Genehmigungsvorgaben gemacht werden, die der Weitergabe entgegenstehen. Erst wenn eine Genehmigung vorliegt, kann mit den unmittelbaren Beteiligten, ihren Erziehungsberechtigten und/oder rechtlichen Vertretungen Kontakt aufgenommen werden.

Das Einverständnis der unmittelbar am Forschungsprozess Beteiligten, ihrer Erziehungsberechtigten und/oder rechtlichen Vertretungen zielt zunächst auf die Möglichkeit der Durchführung der Primärforschung. Darüber hinaus ist mit den Beforschten, ihren Erziehungsberechtigten und/oder rechtlichen Vertretungen zu klären, ob und inwieweit Formen der Sekundärnutzung möglich sind. In jedem Fall sollten Beforschte aber über Ort und Dauer der Datenspeicherung (z. B. auch im Fall einer Archivierung für min. 10 Jahre bei den Primärforschenden; siehe Abschnitt 4.) informiert werden.

Die Bereitschaft zur Nutzung durch weitere Forschende darf nicht zulasten der Zustimmung zum Primärforschungsvorhaben gehen. Dieser Aspekt ist gleichermaßen aus Sicht der Beforschten als auch aus forschungsethischer Sicht zu prüfen. Daher können, wenn zu erwarten ist, dass Beforschte einer Sekundärnutzung der Daten kritisch gegenüberstehen, das Einverständnis zur Primärstudie und das Einverständnis zur Bereitstellung der Forschungsdaten separat abgefragt werden. Dies gilt insbesondere vor dem Hintergrund der Vulnerabilität von Erziehungs- und Bildungsprozessen bzw. deren Erforschung. Die Beforschten, ihre Erziehungsberechtigten und/oder rechtlichen Vertretungen sind darauf hinzuweisen, dass ihnen durch die Nicht-Teilnahme an dem Forschungsvorhaben keinerlei Nachteile entstehen und dass eine Einverständniserklärung jeder Zeit und ohne Angabe von Gründen zurückgezogen werden kann.

<sup>10</sup> [https://www.ratswd.de/dl/RatSWD\\_Output3\\_Forschungsdatenmanagement.pdf](https://www.ratswd.de/dl/RatSWD_Output3_Forschungsdatenmanagement.pdf) letzter Zugriff: 11.03.2020

In der Information über das Projekt und den Umgang mit den Daten gegenüber den Beforschten, ihren Erziehungsberechtigten und/oder rechtlichen Vertretungen sind Angaben zu machen, wie ihre Datenschutzrechte in allen Forschungsphasen berücksichtigt werden, u.a. wie Anonymisierungen und Pseudonymisierungen vorgenommen werden, im Rahmen der Bearbeitung der Daten sowie ihrer Präsentation in Form von Publikationen, Lehrveranstaltungen und/oder wissenschaftlichen Vorträgen.

Die Primärforschenden müssen abwägen und ihre Überlegungen begründend ausführen, inwiefern durch die Anforderung der Archivierung von Daten zum Zweck der Nachnutzung das primäre Forschungsziel gefährdet wird. In solch einem Fall darf der Umstand, dass die Beforschten bei der Primärerhebung ihr Einverständnis in die Bereitstellung ihrer Daten nicht geben, nicht dazu führen, dass insgesamt keine Forschung durchgeführt werden kann. In vergleichbarer Weise sind auch Fragen des Urheberrechts und diesbezügliche ethische Fragen zu behandeln. Weder die Rechte der Primärforschenden, noch die Verpflichtungen die sie eingegangen sind, dürfen durch die Bereitstellung von Forschungsdaten übergangen werden.

Im Fall der Absicht, (aufbereitete) Daten mit dem Zweck der Ermöglichung einer Sekundärnutzung zu archivieren, ist dies in der Einverständniserklärung entsprechend aufzuführen. Mit Zustimmung der Beteiligten können auch Daten mit Personenbezug wie etwa nicht-anonymisierte Rohdaten, wie z.B. Videos oder Audiodateien (z.B. Oral History) archiviert und nachgenutzt werden. Grundlegend gilt, dass eine einmal gegebene Einverständniserklärung jederzeit für die Beforschten, ihre Erziehungsberechtigten und/oder rechtlichen Vertretungen widerrufen werden kann. Je nach Forschungsvorhaben und -vorgehen kann ausgewiesen werden, dass nur ein Teil der Daten archiviert und weitergegeben wird, andere hingegen nicht. Dies ist in den Einverständniserklärungen zu berücksichtigen.

Bei geplanter Sekundärnutzung ist in der Einverständniserklärung zu explizieren, wie die geplante Weitergabe der aufbereiteten Daten (in pseudonymisierter oder anonymisierter Form) sowie der Kontextdaten/Kontextinformationen zur Nachnutzung erfolgen soll. Bei der Erstellung von Einverständniserklärungen kann auf die Expertise der Forschungsdatenzentren (FDZ) oder der lokalen Ethikkommissionen zugegriffen werden<sup>11</sup>.

### **3.3 Während des Projektes: Datenerhebung und Datenaufbereitung**

Datenerhebung bedeutet die Generierung von Rohdaten. Diese Rohdaten werden meist am Schnittpunkt von Erhebung und Auswertung zum Zweck der Auswertung aufbereitet; aus Rohdaten entstehen aufbereitete Daten. In der Regel kommen für die Archivierung und Nachnutzung von Forschungsdaten nur die aufbereiteten Daten in Frage, d.h. diejenigen, die für die Auswertung im Rahmen der Primärforschung aufbereitet wurden, z.B. nur transkribierte Passagen aus längeren Interviews. Relevant in Bezug auf die Archivierung sind dabei (a) technische, (b) datenschutz- und urheberrechtliche sowie (c) forschungsethische Aspekte. Im Datenmanagementplan ist jeder dieser Aspekte zu beachten/auszuweisen.

- a) Technische Aspekte der Veränderung der ‚Struktur‘ des Datenmediums im Zuge der Aufbereitung

Im Zuge der Aufbereitung von Daten werden Rohdaten i.d.R. so verarbeitet, dass sie zur Auswertung zur Verfügung stehen: So entstehen (codierte) Transkripte bspw. aus Audio- und Videodaten oder werden Antwortdaten aus Tests und Fragebögen in pseudonymisierte/anonymisierte Datensätze umgewandelt. In diesem Prozess werden Dokumentationsmaterialien – Kodierleitfäden,

---

<sup>11</sup> Siehe z. B. <https://www.forschungsdaten-bildung.de/einwilligung>; letzter Zugriff: 11.03.2020

Pseudonymisierungs/Anonymisierungsregeln – erstellt, die als (Teil der) Kontextdaten/Kontextinformationen selbst wiederum Teil der Generierung von Daten sind. Diese Kontextdaten/Kontextinformationen sollen es auch Forschenden, die an der Primärerhebung nicht beteiligt waren, ermöglichen, die Erhebung und Aufbereitung der Daten nachzuvollziehen und mit den Daten zu arbeiten. Diese Kontextdaten sind datenschutzrechtlich und forschungsethisch hoch sensibel. Insofern ist ihre Archivierung und Zugänglichmachung mit Blick auf die Bereitstellung gesondert zu betrachten (siehe Punkt 4.1).

Im Besonderen bei qualitativen Forschungsdaten ist es zum Teil im Zuge der Aufbereitung der Rohdaten nicht möglich, die Daten so zu pseudonymisieren / anonymisieren, dass sie ihren Informationsgehalt nicht verlieren. Dies trifft etwa auf bildförmige und videografische Daten oder Beobachtungsprotokolle zu. So können z.B. Bild- und Videodaten schon aufgrund der technischen Möglichkeiten der Bilderkennung nicht vollständig als pseudonymisiert gelten. Zwar ist eine Weitergabe von Bild- und Videodaten weitgehend pseudonymisiert durch Pixeln der Gesichter/Körper möglich, allerdings fehlen so u.U. auch einige für die Nachnutzung relevante Informationen. Insofern können Primärforschende zwar versuchen, Genehmigungen und Einverständniserklärungen für die Nachnutzung dieser teil-pseudonymisierten Bild- und Videodaten zu erhalten (z.B. ohne Namensnennung, doch ohne Pixelung), die mit einem (gesicherten) Zugang per Antrag einhergeht (vgl. Kasten 3). Sollten allerdings Beforschte bzw. deren rechtliche Vertretungen oder die Genehmigungsregelungen der Träger/Behörden dies nicht erlauben, dürfen nur anonymisierte Transkripte zur Nachnutzung bereitgestellt werden. Diese Praxis sollte im Vergabe- und Begutachtungsverfahren für Projekte nicht nachteilig ausgelegt werden.<sup>12</sup>

Bezüglich des Aufbereitungsstatus von quantitativen Daten bietet sich die Archivierung des ersten qualitätsgesicherten Datensatzes an, in dem die Daten bereits bereinigt wurden und hilfreiche Aufbereitungsschritte erfolgt sind, aber das Analysepotenzial nicht durch zu starke Vergrößerung von Informationen eingeschränkt ist. Zu diesen Schritten zählen etwa das Bereinigen von Eingabefehlern, das Entfernen doppelter Informationen, das Hinzufügen von Variablenlabeln, eine einheitliche Definition und Kodierung fehlender Daten (z.B. auch unter Berücksichtigung der Filterführung in Fragebögen) und die Kodierung von Leistungstests (Scoring). Auch hilfreiche Aggregationen und abgeleitete Variablen sollten enthalten sein (z.B. gebildete Indices für den sozialen Hintergrund, Skalenwerte oder, wenn verfügbar, IRT-basierte Personenschätzer für Leistungstests). Diese Aggregationen sollten aber nicht so weit gehen, dass eigene Aufbereitungen für Zusatzanalysen der nachnutzenden Forscherinnen und Forscher nicht mehr möglich sind. So sollte z.B. die Einzelitem-Ebene in Datensätzen enthalten sein (und nicht etwa nur Skalenwerte oder Ergebnisse aus IRT-Skalierungen). Zusätzlich sollten die Schritte vom Rohdatensatz zu dieser Datensatzversion im Dokumentationsmaterial nachvollziehbar dargestellt werden. Es kann auch sinnvoll sein, mehrere Datensätze in verschiedenen Aufbereitungsgraden oder mehrere Versionen der gleichen Variable (z. B. roh/imputiert, wenn multiple Imputation fehlender Werte im Rahmen des Primärprojekts zur Anwendung kam) bereitzustellen. Weitere Hinweise zur Datenaufbereitung finden sich zum Beispiel beim VerbundFDB<sup>13</sup>. Darüber hinaus gibt es aktuell eine Initiative von Science Europe, die darauf abzielt, sogenannte Domain Data Protocols (DDPs) zu entwickeln – standardisierte Datenmanagementpläne für verschiedene Fachbereiche, die Forschenden helfen sollen, qualitätsgesicherte Forschungsdaten nach aktuellen Standards zu produzieren<sup>14</sup>. Somit ist in Zukunft mit einer verstärkten Standardisierung von Datensatzformaten und Aufbereitungsstandards und darauf

---

<sup>13</sup> Trixa, Jessica und Thomas Ebel. 2015. Hinweise zur Aufbereitung quantitativer Daten. *forschungsdaten bildung informiert*, Nr. 4. Online verfügbar unter: <https://www.forschungsdaten-bildung.de/publikationsreihen>; letzter Zugriff: 04.07.2018

<sup>14</sup> [https://www.scienceurope.org/wp-content/uploads/2018/01/SE\\_Guidance\\_Document\\_RDMPs.pdf](https://www.scienceurope.org/wp-content/uploads/2018/01/SE_Guidance_Document_RDMPs.pdf); letzter Zugriff: 11.03.2020

basierenden Handreichungen und Tools zu rechnen, die Primärforscherinnen und Primärforschern das Datenmanagement erleichtern wird.

#### b) Datenschutz- und eigentümer-/urheberrechtliche Aspekte

In Abhängigkeit von den Personen und sozialen Gruppen, den gewählten Erhebungsmethoden und den Auswertungsverfahren stellen sich die datenschutz- und eigentümer-/urheberrechtlichen Anforderungen der Archivierung von Forschungsdaten sehr unterschiedlich dar. Während in einigen Fällen mit vergleichsweise geringem Aufwand eine Deanonymisierung /Depseudonymisierung der Forschungsdaten praktisch ausgeschlossen werden kann, ist es in anderen Fällen aufgrund der Charakteristik der untersuchten Gruppen und/oder des Umfangs der erhobenen Informationen über einzelne Personen nicht möglich, die Daten zu anonymisieren, ohne dass sie ihren Informationsgehalt verlieren. Zudem sind „Forschende [...] in unterschiedlichem Maße, in der qualitativen Forschung jedoch typischerweise sehr ausgeprägt, aktiv in die Produktion von Datenmaterial eingebunden. Insbesondere in ethnographischen Verfahren werden sie zu Autor\*innen ihres Materials, das umgekehrt immer bereits Elemente der analytischen Eigenleistung enthält.“<sup>15</sup> Die Frage der Weitergabe berührt in diesen Fällen also in besonderem Maße die Frage des geistigen Eigentums.“<sup>16</sup>

Datensätze können nur dann ohne Einverständnis der Teilnehmenden (bzw. ihrer Erziehungsberechtigten und/oder rechtlicher Betreuung) weitergegeben werden, wenn die aufbereiteten Daten *keinen Personenbezug* (mehr) aufweisen.<sup>17</sup> Zwar enthalten auch anonymisierte oder pseudonymisierte Daten keine Angaben mehr, mit denen sich ein direkter Personenbezug herstellen lässt, ob diese Daten allerdings im Sinne der geltenden Datenschutzgesetzgebung als gar nicht mehr personenbezogen gelten, ist abhängig von einer Einschätzung, ob und mit welchen Mitteln bzw. welchem Aufwand eine Re-Identifikation einzelner Personen möglich wäre und welcher Schaden den betroffenen Personen bei einer möglichen Re-Identifizierung entstehen würde. Hier ist einzubeziehen, dass technische Möglichkeiten, die dies eröffnen, sich wandeln können und werden.

Im Datenmanagementplan ist insofern aufzuführen, ob und in welcher Form Daten mit direktem Personenbezug gespeichert werden bzw. zu begründen, inwiefern diese nicht gespeichert werden (können). Hierzu empfiehlt sich die Rücksprache mit der bzw. dem Datenschutzbeauftragten der eigenen Organisation.

#### c) Forschungsethische Aspekte

Die Spezifika erziehungs- und bildungswissenschaftlicher sowie fachdidaktischer Forschung erfordern eine besondere Reflexion der forschungsethischen Implikationen, da sie auf die Erhebung von Daten konturiert sind, die sich auf Bildung, Erziehung, Lernen und Entwicklung von bzw. durch Menschen auszeichnen. Dies gilt v.a. für Kinder und Jugendliche, auf die sich ein Großteil der Forschung bezieht, und die im Besonderen eine schützenswerte und vulnerable Gruppe darstellen. Dies gilt aber auch dort, wo – bspw. im Rahmen der Lehrer\*innenbildung – Forschung in (universitären) Bildungskontexten durchgeführt wird. Gerade in Bezug auf Bildung- und Erziehungsprozesse in (staatlichen) Organisationen ist zu reflektieren, dass Praktiken und/oder Äußerungen in

---

<sup>15</sup> So stellt z.B. das Schreiben von ethnografischen Beobachtungsprotokollen, die ja die Datengrundlage für weitere Auswertungen darstellen, bereits eine analytische Eigenleistung der bzw. des Schreibenden dar, die in der „sprachliche[n] Erschließung von Phänomenen [...] d.A.), die noch gar nicht in sprachlicher Form vorliegen“ (Georg Breidenstein, Stefan Hirschauer, Herbert Kalthoff & Boris Nieswand. 2013. *Ethnografie. Die Praxis der Feldforschung*. Wien: UTB. S. 35), liegt. Denn durch die bzw. den Forschende\*n wird im Schreiben eine „vielschichtige soziale Welt nicht nur in eine zweidimensionale Form – die Schrift – übersetzt, sondern erst in Sprache überführt, benannt und bezeichnet“ (ebd.).

<sup>16</sup> Rat SWD (2015): Stellungnahme des RatSWD zur Archivierung und Sekundärnutzung von Daten der qualitativen Sozialforschung, S. 5. Online verfügbar unter: [https://www.ratswd.de/dl/RatSWD\\_Stellungnahme\\_QualiDaten.pdf](https://www.ratswd.de/dl/RatSWD_Stellungnahme_QualiDaten.pdf); letzter Zugriff: 26.11.2019

<sup>17</sup> [https://www.ratswd.de/dl/RatSWD\\_WP\\_264.pdf](https://www.ratswd.de/dl/RatSWD_WP_264.pdf); letzter Zugriff: 11.03.2020

organisatorischen Zusammenhänge eingebunden sind, in denen spezifische Erwartungen an die Rolle der Beteiligten gestellt werden, mit denen auch Abhängigkeiten einhergehen. Insofern sind Aussagen zu den forschungsethischen Implikationen der Archivierung ein Aspekt, die die Eignung bzw. Zurückhaltung von Forschungsdaten für die Archivierung und Nach-/Sekundärnutzung bzw. spezifische Formen der Archivierung und Nachnutzung begründen können. Auch hier empfiehlt sich die Rücksprache bzw. Prüfung durch die/den Datenschutzbeauftragte/n und die Ethikkommission der eigenen Organisation.

**Kasten 2: Beispiele für forschungsethische Aspekte in Bezug auf die Entscheidung für / gegen Datenarchivierung bzw. Nachnutzung**

Beispiel 1: In einer Untersuchung zu Konflikten im familialen Alltag werden Interviewdaten von Kindern zu einem spezifischen Zeitpunkt ihrer Entwicklung festgehalten, in denen diese auf konkrete Personen – Eltern und Geschwister – Bezug nehmen. Ihre Haltung zu diesen Personen zum Zeitpunkt Y wird damit möglicherweise in einer Form festgeschrieben, wie sie für die Kinder aufgrund von Entwicklungs-, Lern-, Sozialisationsprozessen zu einem späteren Zeitpunkt Z nicht mehr gültig sind, und wie sich die Kinder selbst nicht mehr sehen wollen. Unabhängig vom Datenschutz stellt sich hier das Problem, dass Kinder (und die Eltern) in pseudonymisierten Veröffentlichungen oder in durch Nachnutzung entstandenen Kontexten durchaus wissen, dass es hier um sie geht.

Beispiel 2: Videografien von Unterrichts- bzw. schulischen Lehrsituationen angehender Lehrkräfte. Die Erhebung von Daten zielt hier auf ein – nur z.T. pseudonymisierbares – Material, dessen Speicherung und Nachnutzung datenschutzrechtlich z.B. dadurch ermöglicht wird, dass die videografierten Studierenden einer Verwendung (auch teil- bzw. unverbildet) zustimmen. Die im Video festgehaltenen möglichen (noch) entwicklungsbedürftigen Aspekte des Handelns der angehenden Lehrperson zum Zeitpunkt X wird allerdings damit möglicherweise später in Kontexten sichtbar, wie z.B. Weiterbildung, in denen (zukünftige) Kolleg\*innen oder Schulleiter\*innen sitzen, die das (ehemalige) Handeln der Person zum (Ausbildungs-)Zeitpunkt X möglicherweise als ein – über die Zeit – kohärentes (festes) Merkmal ihrer (fehlenden) Fähigkeiten deuten und entsprechend bewerten.

Beispiel 3: Ein Ethnograf beobachtet eine Unterrichtssituation, die von Demütigungen der Schüler und Schülerinnen gegenüber der hilflos erscheinenden Lehrerin geprägt sind.<sup>18</sup> Bereits die Beobachtungssituation erscheint ambivalent, weil der Ethnograf zwischen der „Bewegung im machtförmigen Feld“<sup>19</sup> und der neutralen Beobachtungsposition changiert. Folgt er den Normalitätskonstruktionen der Akteure im Feld, beobachtet er die ‚Verletzungen‘, ohne zu intervenieren. Durch das ‚Festhalten‘ der Demütigungen im Beobachtungsprotokoll entsteht neben den erkenntnistheoretischen Schwierigkeiten die forschungsethische Problematik, dass mit jeder Nachnutzung dieses Materials die ‚Berechtigung‘ dieser machtförmigen Strukturen des Feldes, gerade durch die der Beobachter\*innenhaltung geschuldete „praktische Parteinahme“<sup>20</sup> bestätigt werden. Zugleich stellt sich für die Nachnutzung der Daten die Frage bzw. Anforderung der Reflexion des „latente[n] Mitschuldigwerden [des Forschers, d.A.] an den Ereignissen“<sup>21</sup>, die forschungsethisch nicht unabhängig vom Primärforschenden geschehen kann.<sup>22</sup>

<sup>18</sup> Michael Meier 2019. Spannungsfelder ethnographischer (Schul- und Unterrichts-)Forschung. In: Jule-Marie Lorenzen, Lisa-Marian Schmidt & Dariusz Zifonun (Hrsg.): Methodologien und Methoden der Bildungsforschung. Quantitative und qualitative Verfahren und ihre Verbindungen (S. 45-64). Weinheim und Basel: Beltz Juventa.

<sup>19</sup> Ebd., S. 46

<sup>20</sup> Ebd., S. 60

<sup>21</sup> Ebd.

<sup>22</sup> Meier setzt sich in erkenntnistheoretischer Hinsicht anhand eines von ihm selbst erstellten Beobachtungsprotokolls einer Unterrichtsstunde kritisch mit den Spannungsfeldern ethnographischer Forschung auseinander. Die Thematisierung der eigenen Rolle im Feld, die in seinem

### 3.4 Nach dem Projektende: Datenarchivierung

#### 3.4.1 Archivierungsmöglichkeiten

Für die Archivierung und Veröffentlichung der aufbereiteten Daten und der zugehörigen Dokumentationsmaterialien (Kontextdaten/Kontextinformationen sowie Erhebungsinstrumente) bieten sich nach Projektende mehrere Möglichkeiten: Die Daten können entweder in der eigenen Institution aufbewahrt, in ein Repositorium gestellt (an der eigenen, falls angeboten, oder einer anderen Institution; fachbezogen oder fächerübergreifend) oder an ein Daten kuratierendes Forschungsdatenzentrum übergeben werden (vgl. Kästen 2 & 3).

#### **Kasten 3: Vertrauenswürdige Archivierungsorte in Deutschland im Bereich der/mit Bezug zur Bildungsforschung**

Einen sehr guten Überblick über Repositorien bietet derzeit (Stand Januar 2020) das Meta-Repositorium [re3data.org](https://re3data.org), in welchem bspw. nach Thema, Fachrichtung, Datentyp, Lizenz, Zertifizierung, Sprache gefiltert werden kann.

Der Rat für Sozial- und Wirtschaftsdaten (RatSWD) weist die Datenzentren<sup>23</sup> im Bereich der Sozialwissenschaften und verwandter Disziplinen nach, die nach dessen Kriterien<sup>24</sup> (mindestens ein Datenzugangsweg, Bereitstellung von Dokumentationen zu den Daten, Konzept zur langfristigen Verfügbarmachung der Daten) akkreditiert sind. Er bietet eine Übersicht über Forschungsdatenzentren (FDZ), die für eine Archivierung in Frage kommen (wobei ein substanzieller Teil der akkreditierten Datenzentren nur Daten der eigenen Institution archiviert und anbietet).

Als spezifische Anlaufstelle für die empirische Bildungsforschung kann das Portal des VerbundFDB<sup>25</sup> dienen. Über das Portal können Forschende an zentraler Stelle Daten übermitteln, die an fachspezifische FDZ zur Archivierung und Bereitstellung weitergeleitet werden.

In der empirischen Erziehungs-, Bildungs- und fachdidaktischen Forschung gibt es eine Vielzahl groß angelegter (quantitativer) Datenerhebungsprojekte mit breiter Messung verschiedener Konstrukte, die für eine Vielzahl verschiedener Publikationen genutzt werden. Hier wird empfohlen, nicht (nur) die Daten einzelner Publikationen bereitzustellen, sondern die *Gesamtdaten eines Projekts an einem zentralen Ort zu archivieren, so dass diese konsistent zitiert werden können* (s. Kasten 3). Dies gilt allerdings für Studien mit unterschiedlichen Datentypen nicht ohne Weiteres. Hier kann es sinnvoll sein, verschiedene Datentypen (z. B. Fragebogendaten, Transkripte oder Videographien) an unterschiedlichen FDZ oder in unterschiedlichen Repositorien zu archivieren, die besondere Expertise für diesen Datentyp haben.<sup>26</sup>

Zudem sind mit der Datenarchivierung in Repositorien oder FDZ und der damit verbundenen Weitergabe der Daten an Dritte datenschutz- und urheberrechtliche Fragen verbunden, die Einfluss auf die Zugänglichmachung zu den und die Möglichkeiten der Nachnutzung der Daten haben. Wichtig ist,

---

Beitrag in erkenntnistheoretischer Hinsicht thematisiert wird, lässt sich unter forschungsethischen Aspekten, die im Beitrag nur am Rand thematisiert werden, zugleich als eine Reflexion der ethischen Implikationen des eigenen Handelns lesen, die in diesem Fall mit der eigenen Involviertheit in die Forschung einhergegangen sind.

<sup>23</sup> <https://www.ratswd.de/forschungsdaten/fdz>

<sup>24</sup> alle Kriterien siehe hier: <https://www.ratswd.de/forschungsdaten/akkreditierung>; letzter Zugriff: 04.07.2018

<sup>25</sup> <https://www.forschungsdaten-bildung.de>; letzter Zugriff: 11.03.2020

<sup>26</sup> Meyermann et al., (2017). Der Verbund Forschungsdaten Bildung – Eine Forschungsdateninfrastruktur für die empirische Bildungsforschung. RatSWD Working Paper 266.

dass die aufbereiteten Daten datenschutzkonform zur Verfügung gestellt werden (siehe Punkt 2: Datenerhebung/Datenaufbereitung).

Für die Archivierung von Kontextdaten, die personenbezogene Informationen enthalten, müssen datenschutzrechtlich abgesicherte Wege der Zugänglichmachung entwickelt werden (z.B. Zugang auf Antrag oder gesicherter Zugang; siehe Kasten 4). Prinzipiell sollte die Archivierung solcher personenbezogenen Kontextinformationen sparsam sein. Das heißt, es sollte grundlegend abgewogen werden, welche Kontextinformationen und in welchem Umfang diese unbedingt archiviert werden müssen, damit eine Nachnutzung der aufbereiteten Forschungsdaten überhaupt sinnvoll möglich ist.

**Kasten 4: Derzeitige Möglichkeiten des Datenzugangs zur Nachnutzung**

*Freier Zugang* („Public Use File“): Daten können, entweder direkt oder nach Registrierung und Zustimmung zu Nutzungsbedingungen, eingesehen oder heruntergeladen werden.

*Zugang auf Antrag* („Scientific Use File“): Daten können erst nach Stellung eines Antrags und auf Basis einer Datennutzungsvereinbarung genutzt werden. Der Datenzugang ist auf wissenschaftliche Nutzung beschränkt. Scientific Use Files sind typischerweise weniger stark anonymisiert als Public Use Files.

*Gesicherter Zugang*: Um Zugang zu Datensätzen mit niedrigerem Anonymisierungsniveau anbieten zu können, bieten einige FDZ geschützte Zugänge an. Dazu zählt die Möglichkeit, Auswertungen an einem Gastarbeitsplatz durchzuführen oder – bei quantitativen Daten – über ein Fernrechner-System (z.B. RemoteNEPS, JoSuA) zu arbeiten. Hierbei erfolgt keine physische Übergabe des Datensatzes an die Nutzerinnen und Nutzer.

Zu beachten bei der Datenübergabe an FDZ ist zudem, dass eine Übertragung von Nutzungsrechten an das Repositorium oder FDZ erfolgt, die über einen Vertrag oder die Zustimmung zu Nutzungsbedingungen geregelt wird. Je nach Archivierungsort und -form können auch Kosten entstehen, die in Projektanträgen mit beantragt werden sollten. FDZ beraten zum Thema Pseudonymisierung bzw. Anonymisierung und wenden ggf. weitere Anonymisierungs- und Pseudonymisierungsstrategien an<sup>27</sup>.

### 3.4.2 Übergabe von Daten zur Archivierung

Die Archivierung von Forschungsdaten dient zum einen der Sicherung und Dokumentation eines Forschungsprojekts für die Primärforschenden selbst (bzw. deren Institution/Arbeitsgruppe), zum anderen aber auch der Nachnutzbarkeit der Daten durch andere Wissenschaftler\*innen. Im Sinne einer offenen und transparenten Wissenschaftskultur ist die Ermöglichung eines Nachvollzugs der generierten Ergebnisse bzw. einer Reanalyse bzw. Replikation von auf Basis der Datensätze publizierten Analysen möglich und unterstützenswert. Wie in den Leitlinien zur Sicherung guter wissenschaftlicher Praxis der DFG empfohlen, sollten Forschungsdaten „in der Regel für einen Zeitraum von zehn Jahren zugänglich und nachvollziehbar in der Einrichtung, wo sie entstanden sind, oder in standortübergreifenden Repositorien aufbewahrt“<sup>28</sup> werden. Für den Fall, dass nachvollziehbare Gründe gegen eine Archivierung bestimmter Daten sprechen, soll dies von den Forschenden erläutert werden. Wenn eine Archivierung in einem (öffentlichen bzw. teilöffentlichen)

<sup>27</sup> Ebel, Thomas und Alexia Meyermann. 2015. [Hinweise zur Anonymisierung von quantitativen Daten](#). *forschungsdaten bildung informiert*, Nr. 3 und Meyermann, Alexia und Maïke Porzelt. 2014. [Hinweise zur Anonymisierung von qualitativen Daten](#). *forschungsdaten bildung informiert*, Nr. 1. Online verfügbar unter <https://www.forschungsdaten-bildung.de/publikationsreihen/>; letzter Zugriff: 04.07.2018

<sup>28</sup> [https://www.dfg.de/download/pdf/foerderung/rechtliche\\_rahmenbedingungen/gute\\_wissenschaftliche\\_praxis/kodex\\_gwp.pdf](https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf), S. 22

Repositoryum oder einem FDZ aus Gründen der Forschungsethik oder datenschutzrechtlichen Gründen nicht erfolgen kann, sollten die Daten an anderer Stelle mindestens 10 Jahre gesichert werden.<sup>29</sup>

Daten können an ein Repositoryum, ein FDZ oder an ein Archiv übergeben werden, sobald diese aufbereitet sind und die Aufbereitung entsprechend dokumentiert wurde. In der Regel sollten Daten mit Projektende übergeben werden. Dabei kann die Übergabe begründet mit einer möglichen Nutzungssperre bzw. einem Embargo für weitere Bearbeitungsmöglichkeiten der Primärforschenden erfolgen. So können etwa zum Schutz von Qualifikationsarbeiten konkrete Forschungsfragen benannt werden, die noch bearbeitet werden (sogenannte Sperrvermerke). Nutzungssperren sind konkret und nachvollziehbar zu begründen und der zeitliche Rahmen der Sperrung ist anzugeben. Das Veröffentlichungsdatum kann in einem Repositoryum eingestellt oder mit einem FDZ vereinbart werden bzw. den archivischen Schutzfristregelungen unterliegen<sup>30</sup> Bei Projekten mit langer Laufzeit, in denen mehrfach Daten erhoben werden (z. B. mehrere Wellen einer Längsschnittstudie oder mehrere Experimente mit längerem zeitlichen Abstand), können die Datensätze einzeln bereitgestellt werden. Hier sollte im Einzelfall entschieden werden, ob eine Bereitstellung von Teildatensätzen oder eine Bereitstellung aller Daten nach Projektende sinnvoller erscheint. Z. B. könnten, je länger die Gesamtlaufzeit der Studie und der Abstand zwischen den Wellen ist, die Daten der jeweiligen Erhebungswellen einzeln bereitgestellt werden. Davon unbetroffen bleibt die unmittelbare Ermöglichung von Replikationen publizierter Analysen, auch wenn die Gesamtdaten einer Studie noch nicht zur Nachnutzung bereitgestellt sind. Die Primärforschenden sollten auf Anfrage Informationen über die Nachnutzung ihrer Datensätze von den FDZ erhalten (z.B. Anzahl der Anträge, Themen).

Vor dem Hintergrund dieser Ausführungen stellen sich Fragen nach den Rechten der Primärforschenden, der Wahrung ihrer Forschungsinteressen und, damit verbunden, nach den Zugangsbedingungen für mögliche Nachnutzung, siehe Kasten 4. Hier ist zunächst die Zitierpflicht der Daten zu nennen, der Sekundärforschende unterliegen, sowie die Darlegung der Weiterver- und -bearbeitung der Daten (s. Kasten 5).

---

<sup>29</sup>[https://www.dfg.de/download/pdf/dfg\\_im\\_profil/reden\\_stellungnahmen/download/empfehlung\\_wiss\\_praxis\\_1310.pdf](https://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf); letzter Zugriff: 11.03.2020

<sup>30</sup> Bei den Forschungsdatenzentren des VerbundFDB und insbesondere bei den Studien des nationalen und internationalen Bildungsmonitorings (PISA, IGLU, TIMSS, IQB-Bildungstrends) hat sich z.B. eine Laufzeit von zwei Jahren nach der erstmaligen Datenveröffentlichung für Sperrvermerke zu konkreten Forschungsfragen etabliert.<sup>31</sup>  
[http://www.forschungsdaten.org/index.php/Persistent\\_Identifier](http://www.forschungsdaten.org/index.php/Persistent_Identifier); letzter Zugriff: 11.03.2020

#### Kasten 5: Möglichkeiten der Zitation von Datensätzen

Die Zitationsrichtlinien der American Psychological Association (APA) erlauben die Zitation von Datensätzen im Literaturverzeichnis in folgender Form:

Namen der Autor\*innen (Jahr). *Datensatzname* [data set]. doi:.

Also beispielsweise:

U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies. (2013). *Treatment episode data set -- discharges (TEDS-D) -- concatenated, 2006 to 2009* [data set]. doi:10.3886/ICPSR30122.v2

Analog könnte für den deutschsprachigen Raum folgende Form verwendet werden, die vom VerbundFDB umgesetzt wird:

Wuttke, E. & Seifried, J. (2018): Diagnose von und Umgang mit Schülerfehlern als Facette der professionellen Kompetenz von Lehrkräften (ProFel II). Version: 1. IQB – Institut zur Qualitätsentwicklung im Bildungswesen. Datensatz. doi:10.5159/IQB\_ProFel\_II\_v1

Ebenso können Begleitmaterialien wie etwa Skalenhandbücher zitiert werden. Bei der Übergabe an ein FDZ oder Repositorium besteht die Möglichkeit, festzulegen, wie die Datensatzzitation aussehen soll (also z. B. Reihenfolge der Autor\*innen) und welche Zusatzdokumente Sekundärforschende zitieren sollen. In jedem Fall sollte der Datensatz mit einem Persistent Identifier (PID<sup>31</sup>; bspw. DOI (Digital Object Identifier)) versehen werden, der die eindeutige und dauerhafte Referenzierbarkeit der Daten gewährleistet.

Eine Nachnutzung erhobener Primärdaten ist nicht in jedem Fall möglich, insbesondere dann, wenn die Anonymität sowohl der Befragten als auch der Forschenden nicht zu gewährleisten ist oder urheberrechtliche Gründe dagegensprechen. Insofern kann die Nachnutzung von Forschungsdaten nicht als generelle Forderung an Forschungsprojekte postuliert werden. Es muss die Möglichkeit bestehen bleiben, Forschungsdaten nicht bzw. nur eingeschränkt zur Verfügung zu stellen. Zu schützen sind hier nicht nur die Beforschten, sondern auch die Forschenden, die, insbesondere in einigen qualitativen Forschungsprozessen, als konkrete Personen in die Forschung involviert und so auch zu „Autor\*innen ihres Materials“<sup>32</sup> werden und Informationen zum Forschungsprozess beitragen. Das Prinzip der Nichtschädigung muss von daher auch für Forschende gelten und diesen muss ein Vetorecht zur Verfügung stehen, wenn sie mit der Publikation der Daten persönliche Informationen weitergeben würden, die diesem Prinzip entgegenstehen. Zu beachten ist jedoch, dass eine Entscheidung gegen eine Nachnutzung der Daten Forschende nicht von der Verantwortung entbindet, die Daten angemessen zu archivieren.

Hinsichtlich der Weitergabe von Daten ergeben sich somit aufgrund des sehr unterschiedlichen Charakters der Daten zusammenfassend mehrere Möglichkeiten:

Möglichkeit 1): Die Daten können prinzipiell an ein Repositorium oder FDZ übergeben werden. Die Weitergabe der aufbereiteten Daten erfolgt spätestens nach Projektende. Spätestens zwei Jahre nach Übergabe (Embargo) stehen die Daten für eine Nachnutzung zur Verfügung, so diese nicht mit einem Sperrvermerk (z.B. für bestimmte Fragestellungen) versehen wurden. Weitere Verlängerungen der

<sup>31</sup> [http://www.forschungsdaten.org/index.php/Persistent\\_Identifier](http://www.forschungsdaten.org/index.php/Persistent_Identifier); letzter Zugriff: 11.03.2020

<sup>32</sup> Rat SWD (2015): Stellungnahme des RatSWD zur Archivierung und Sekundärnutzung von Daten der qualitativen Sozialforschung, S. 5. Online verfügbar unter: [https://www.ratswd.de/dl/RatSWD\\_Stellungnahme\\_QualiDaten.pdf](https://www.ratswd.de/dl/RatSWD_Stellungnahme_QualiDaten.pdf); letzter Zugriff: 26.11.2019

<sup>32</sup> [https://www.ratswd.de/dl/RatSWD\\_WP\\_264.pdf](https://www.ratswd.de/dl/RatSWD_WP_264.pdf); letzter Zugriff: 11.03.2020

Nutzungssperren bedürfen eines inhaltlich begründeten Antrages. Eine dauerhafte Sperrung bzw. Teilspernung ist im Fall besonders sensibler Daten (siehe Punkt 2) möglich.

Möglichkeit 2: Bei besonders sensiblen personenbezogenen Daten (für Rohdaten wie auch für aufbereitete Daten) ist die Möglichkeit zu prüfen, diese an ein geeignetes öffentliches Archiv zu übergeben. Eine Nutzung der Daten unterliegt dann den jeweils geltenden Archivgesetzen. Für personenbezogene Daten gilt eine Schutzfrist von zehn Jahren nach dem Tod der betroffenen Person. Falls das Todesjahr unbekannt ist, endet die Schutzfrist 100 Jahre nach der Geburt der Person. Falls weder Todes- noch Geburtsdatum bekannt oder nur mit unverhältnismäßigem Aufwand festzustellen sind, so endet die Schutzfrist 60 Jahre nach Entstehung der Unterlagen.<sup>33</sup> Eine Übergabe der Daten an ein öffentliches Archiv würde es ermöglichen, eine Vielzahl von Roh-/ und aufbereiteten Daten für die (bildungs-)historische Forschung zu erhalten.

Möglichkeit 3: In Fällen, in denen das Prinzip der Nichtschädigung für Beforschte und Forschende ethisch und datenschutzrechtlich nicht gesichert ist bzw. mit hoher Wahrscheinlichkeit nicht gesichert werden kann, muss die Möglichkeit bestehen, auf eine Sekundärnutzung zu verzichten. Hierauf ist bereits im Forschungsdatenmanagementplan hinzuweisen.

Der Ablauf des Bereitstellungsprozesses von Forschungsdaten ist überblicksartig in Abbildung 1 dargestellt.

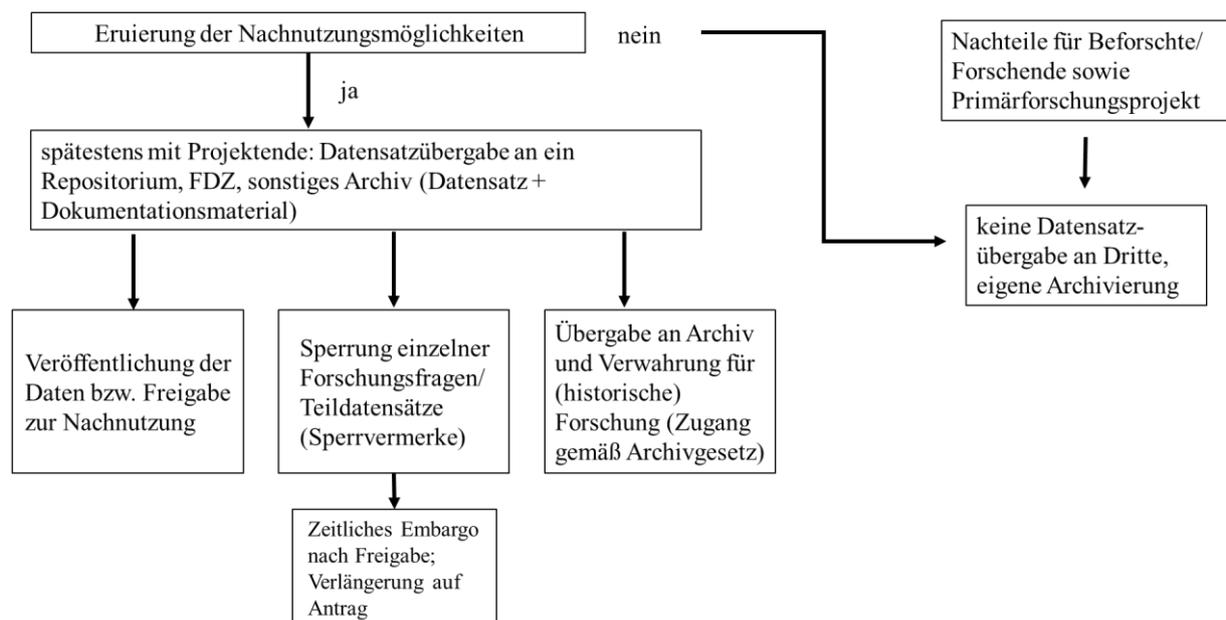


Abbildung 1: Übersicht über Ablauf des Datenbereitstellungsprozesses

<sup>33</sup> Vgl. Hessisches Archivgesetz §13 Schutzfristen bzw. vergleichbar in weiteren Landesarchivgesetzen.

## **4 Notwendige Rahmenbedingungen für die Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten – Empfehlungen an Bildungs- und Forschungspolitik**

Die dargelegten „Empfehlungen zur Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten im Kontext erziehungs- und bildungswissenschaftlicher sowie fachdidaktischer Forschung“ der drei Fachgesellschaften richten sich insbesondere an die Wissenschaftler\*innen der Fachgesellschaften, die Daten erheben und ggf. zur Nachnutzung zur Verfügung stellen. Damit diese Empfehlungen Praxis werden können, bedarf es aber auch entsprechender Rahmenbedingungen. Folgende Strategien und Maßnahmen sind aus Sicht der drei Fachgesellschaften notwendig, damit eine Kultur des Open Science im Bereich der Datenerhebung, -bereitstellung und -nachnutzung etabliert werden kann:

So ist ein Umdenken in der Bewertung wissenschaftlicher Leistungen notwendig. So sollten Sekundäranalysen in unserem Forschungsgebiet die Wertschätzung erhalten, die sie in einigen Disziplinen bereits besitzen (z. B. der Bildungssoziologie und der Bildungsökonomie). Sekundäranalysen tragen zur effizienten Nutzung von Ressourcen und zur Reduktion der Belastung von Bildungsinstitutionen bei. Sekundäranalytische Projekte sollten ebenso gute Chancen auf Fördermittel haben wie Primärerhebungen. In dieser Hinsicht sind Initiativen wie die Förderung der Nachwuchsakademie „Sekundäranalysen multidisziplinär nutzbarer Datensätze der Bildungsforschung“ durch die DFG zu begrüßen. Nicht nur bei Forschungsprojekten, sondern auch bei der Betreuung von Qualifikationsarbeiten sollte die Möglichkeit der Sekundärdatenanalyse einer Neuerhebung von Daten gegenübergestellt werden. Es sei aber darauf hingewiesen, dass Sekundäranalysen nur dann fruchtbare Beiträge leisten können, wenn eine hohe Passung zwischen Forschungsfragen (und einem hypothetischen, für die Beantwortung der Fragen idealen Datensatz) und den tatsächlich im Datensatz erfassten Merkmalen bzw. für die Datenerhebung genutzten Messinstrumenten besteht. Eine erhöhte Möglichkeit und Wertschätzung von Sekundärdatenanalysen soll aber in keiner Weise mit einer geringeren Wertschätzung von Datenproduktion einhergehen, im Gegenteil:

Weiterhin muss für die Primärforschenden die Produktion von Daten, die dann von anderen Wissenschaftlerinnen und Wissenschaftlern nachgenutzt werden, zukünftig stärker als wissenschaftliche Leistung wahrgenommen werden. Die Produktion qualitativ hochwertiger Datensätze sollte ähnlich wie Publikationen oder eingeworbene Drittmittel bei der Bewertung wissenschaftlicher Lebensläufe und daraus folgend bei der Ausschreibung und Besetzung von Stellen berücksichtigt werden. Dafür ist unter anderem die eindeutige Zitierbarkeit von Datensätzen notwendig sowie Zitationspflichten von Nachnutzenden (s. Kasten 5). Auch die Präsentation von Forschungsdatensätzen in eigenen Fachzeitschriften (*data journals*) oder Rubriken von Fachzeitschriften stellt eine Möglichkeit dar, Datensatzproduktion durch eine Publikation angemessen zu würdigen. Es bedarf jedoch unbedingt einer Infrastruktur zur Archivierung und Bereitstellung von Forschungsdaten. Dies betrifft zum einen die Arbeit der verschiedenen FDZ, Repositorien und weiterer Archive, die bereits wichtige Anlaufstellen für die Forschenden sind und zentrale Aufgaben der Datenarchivierung, Bereitstellung und Katalogisierung erfüllen. Weiter sind die Universitäten und Hochschulen in der Pflicht, lokal fächerspezifische Ethikkommissionen und Beauftragte für Datenarchivierung zu etablieren, die mit der/dem Datenschutzbeauftragten der Universität zusammenarbeiten. Hierfür sind entsprechende Mittel bereitzustellen.

Sowohl die FDZ als auch die lokalen Anlaufstellen sollten in der Lage sein, Forschenden professionelle Beratung zu Aspekten des Datenschutzes und damit verbundener forschungsethischer

Fragen zu bieten und Vorlagen für Genehmigungsverfahren zu liefern. Für die Forschenden müssen *rechtskonforme Standards, Handreichungen und Schulungen zum Forschungsdatenmanagement* bereitgestellt und kontinuierlich weiterentwickelt werden. Die Fachgesellschaften unterstützen daher Bestrebungen zum weiteren Ausbau der Forschungsdateninfrastruktur in Deutschland. Im Zusammenhang damit ist zu beachten, dass die Aufbereitung von Forschungsdaten oft die im Primärprojekt selbst nicht (vollständig) aufbereiteten und/oder ausgewerteten Daten betrifft. Die Umwandlung dieser Rohdaten in – für die Archivierung taugliche – aufbereitete Daten und Datensätze stellt insofern einen eigenen Arbeitsschritt dar, der personeller und finanzieller Ressourcen bedarf, die bei der Forschungsförderung angemessen zu berücksichtigen sind.

Auch wenn die Gesellschaften sich grundsätzlich für eine Kultur der Open Science aussprechen, so darf *im Rahmen der Begutachtung von Forschungsprojekten* die Genehmigung nicht davon abhängig gemacht werden, ob nachnutzbare Daten im geplanten Projekt entstehen. Die Nachnutzung von Forschungsdaten kann nicht als generelle Anforderung an die Projektierung von Forschungsvorhaben formuliert werden. Überlegungen zu Archivierung und Bereitstellung sollten jedoch Teil von Forschungsanträgen sein und die begründete Auseinandersetzung mit der Thematik kann als ein Qualitätskriterium für die Begutachtung herangezogen werden.

Die Bereitstellung von Daten zur Nachnutzung darf nicht dazu führen, dass der *Zugang zum Feld* unserer Forschung, zukünftig (weiter) erschwert wird. Bereits jetzt zeigen intransparente und restriktive Genehmigungsverfahren sowie fehlende administrative Unterstützung dazu, dass viele Forschungsvorhaben speziell an öffentlichen Einrichtungen, wie Schulen, nur sehr begrenzt und häufig unter schwer umzusetzenden Vorgaben realisiert werden können. Dies führt zu deutlichen Einschränkungen der Forschungsqualität. Wir wünschen uns von Seiten der Bildungspolitik eine forschungszugewandte Haltung, die Forschung an Bildungs- und Erziehungseinrichtungen nicht als Eingriff, sondern Teil des pädagogischen Alltags versteht und Forschungsvorhaben daher offen unterstützt. Die von den Fachgesellschaften unterstützte Kultur der Open Science kann zu einer Entlastung der zu beforschenden Institutionen führen, ohne als Argument für einen noch stärker eingegrenzten Feldzugang angeführt werden.

Im Sinne einer offenen Wissenschaftskultur, in der Sekundäranalysen von Datensätzen eine wichtige Rolle spielen, muss auch über Möglichkeiten nachgedacht werden, Daten der Bildungsverwaltung (z.B. Schulstatistiken, Inspektionsberichte) besser zugänglich und damit wissenschaftlich nutzbar zu machen. Dies sollte auch eine bessere Verknüpfbarkeit von Schulstatistik und anderen Forschungsdaten beinhalten.

Schließlich empfehlen wir eine begleitende *Implementationsforschung* zum Umgang mit Forschungsdaten, die die Potenziale und Grenzen, Herausforderungen bei der Umsetzung, Veränderungen in der Wissenschaftskultur u. ä. wissenschaftlich untersucht. Hierzu sollten beispielsweise die FDZ zusätzliche Mittel erhalten oder entsprechende Förderschwerpunkte eingerichtet werden.